

医療統計のための

# データクリーニング実践マニュアル

エラーを起こさず、効率よく行うデータの入力・整理・処理  
ExcelとStataを使って解説

## 目次

■ 序	道端伸明, 麻生将太郎, 藤雄木亨真	3
■ サンプルデータおよびRコード ダウンロードのご案内		8

## 講義編

第1章 臨床研究とは	10
第2章 データクリーニングとは	14
第3章 データ収集に利用できるソフトウェア・Webサービス	16
第4章 統計解析ソフト	20

## 実践編①：データ収集

### Excel ～基本

第1章 Excelの基本操作	24
第2章 データシート管理の基本	31
第3章 変数名入力ルール	36
第4章 データの入力ルール	39
第5章 定義書	42

## Excel ～アドバンス

第6章 入力規制	45
◆コラム①◆ Excelでセルの値と文字列を連結・同じ処理をする	53
◆コラム②◆ Excelで同じ記号を使ってセルの中身を分割	56
◆コラム③◆ Excelの重複の削除	58

## 実践編②：データクリーニング

実践編②の準備, 注意事項	62
---------------	----

## 統計解析ソフト ～基本

第1章 Stataの概要	65
第2章 doファイルとlogファイル	71
第3章 ファイルの種類, パスとディレクトリ	76
第4章 データの読み込み, 保存	81
◆コラム④◆ Unicodeとは? ～文字コードとStataでの注意点	92
第5章 データの確認	94
第6章 変数を作成する, 削除する	108
◆コラム⑤◆ 欠損値について	123
第7章 変数名と変数の値にラベルを付ける	125
第8章 高度な変数作成	127
◆コラム⑥◆ rowmin/rowmaxとrowfirst/rowlastとの違い	141
第9章 グループ毎に集計, 新しい変数を作成する: bysort	143

## 統計解析ソフト ～中級

---

第10章	日付データ	150
◆コラム⑦◆	Storage type	160
第11章	時刻データ	162
◆コラム⑧◆	コンピュータにとっての日付	171
第12章	文字列データ	173
第13章	高度な文字列操作	185
◆コラム⑨◆	正規表現について	199
第14章	フォーマット (format) について	202
第15章	解析対象を制限する	206
第16章	重複データの取り扱い	209
◆コラム⑩◆	パネルデータ解析 ～繰り返し測定データの分析方法	219
第17章	縦持ちデータと横持ちデータの変換	224
第18章	複数のデータセットを統合する, 追加する	232
第19章	ado ファイルのインストール	249

## 統計解析ソフト ～アドバンス

---

第20章	マウス操作	258
第21章	エラー	270
第22章	マクロ機能と local	276
第23章	繰り返しの処理	281
第24章	結果の格納と変数リスト (varlist・ds)	286
第25章	乱数の作成	299

◆コラム⑪◆ Stataの乱数 ～「たかが乱数、されど乱数」	308
第26章 独自プログラムの作成	310
第27章 シミュレーションデータの使用, 作成	325
◆コラム⑫◆ Mataとは	341
第28章 ヘルプ・さらに使いこなすために	344
■ 索引	348
■ あとがき	道端伸明, 麻生将太郎, 藤雄木亨真 350
■ 著者プロフィール	351

### ■ アイコンの説明

 主にマウス操作の説明

 主にキーボードでのコマンド入力の説明